DOCUMENT RESUME

ED 183 608 TM 800 014

AUTHOR Tatsuoka, Kikumi; Birenbaum, Menucha
TITLE The Danger of Relying Solely on Diagnostic Adaptive Testing When Prior and Subsequent Instructional Methods Are Different. Research Report No. 2, October 1 through December 31, 1979.
INSTITUTION Illinois Univ., Urbana. Computer-Based Education Lab.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE Mar 79
CONTRACT N00014-78-C-0159
NOTE 33p.; For related document, see TM 800 036

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Arithmetic; Cluster Analysis; *Computer Assisted Instruction; *Diagnostic Tests; Discriminant Analysis; Instructional Materials; Item Analysis; Junior High Schools; Mathematics; Response Style (Tests); Statistical Analysis; *Student Testing; *Teaching Techniques; *Testing Problems
IDENTIFIERS *Adaptive Testing; *Computer Assisted Testing; Latent Trait Theory; PLATO

ABSTRACT

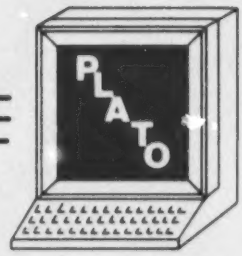
A situation in which computer-assisted instruction was combined with teacher-taught classes and computerized diagnostic adaptive testing was examined. The computerized diagnostic adaptive test for pre-algebra signed-number lessons was programmed along with a computer-managed routing system by which examinees were sent to the instructional unit corresponding to the skill level at which they stopped in the initial test. Upon completing the course a computerized conventional post test was administered. It was found that results differed markedly from results obtained in a previous study. A cluster analysis was performed on the response patterns of the skills, and four different groups were found. A discriminant analysis indicated significant differences among the four groups in response patterns of the skills in signed number operations. It was determined that differences between prior and current instructional methods confused students and caused confusion in the post-test data, and that the scoring procedure of the adaptive testing was not sensitive to individual differences in information processing skills which were affected by the instructional method used in previous teaching. It was concluded that reliance should not be placed solely on test results determined by performance scores on a diagnostic pretest. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED183608

Computer-based Education

Research Laboratory



U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

University of Illinois — Urbana Illinois

THE DANGER OF RELYING SOLELY ON DIAGNOSTIC ADAPTIVE TESTING WHEN PRIOR AND SUBSEQUENT INSTRUCTIONAL METHODS ARE DIFFERENT

KIKUMI TATSUOKA

MENUCHA BIRENBAUM

Approved for public release; distribution unlimited. Reproduction in whole or in part permitted for any purpose of the United States Government.

This research was sponsored by the Personnel and Training Research Program, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-78-C-0159. Contract Authority Identification Number NR 150-415.

PEG PLATO® Education Group

CERL REPORT E-5

MARCH 1979

TM800 014

Errata

The following is a corrected version of Table 1 on page 9; means and S.D.s have been corrected.

Table 1

Means and Item Difficulties of the Twelve Skills in the Pretest of Signed Numbers in Spring 1978

	Skill Type	Mean and SD		Difficulty
1	2>-5	.870	.218	-1.67
2	6 right of -2	.814	.294	-1.51
3	(-2)+(-5)	.802	.299	-1.37
4	-8+7	.787	.320	-1.06
5	2+(-6)	.750	.377	-.86
6	(-3)-(-2)	.635	.389	-.76
7	5-6	.665	.422	-.49
8	-(-7)	.502	.346	-.01
9	(-4)-(-6)	.467	.439	.10
10	-1-5	.448	.432	.15
11	2-(-7)	.427	.461	.22
12	(-6)-(+5)	.378	.431	.36

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Research Report No. 2
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): The Danger of Relying Solely on Diagnostic Adaptive Testing When Prior and Subsequent Instructional Method are Different

5. TYPE OF REPORT & PERIOD COVERED: 1 October - 31 December 79

6. PERFORMING ORG. REPORT NUMBER: CERL Report E-5

7. AUTHOR(s): Kikumi Tatsuoka, Menucha Birenbaum

8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0159

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Computer-based Education Research Laboratory, University of Illinois, Urbana, IL 61801

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61133N; RR042-04; RR042-04-01; NR150-415

11. CONTROLLING OFFICE NAME AND ADDRESS: Personnel and Training Research Programs, Office of Naval Research (code 458), Arlington, VA 22217

12. REPORT DATE: February 1979

13. NUMBER OF PAGES:

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office):

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number): latent trait theory, logistic model, diagnostic adaptive test, computer managed routing system, information-processing strategies, integers operations, cluster analysis, unidimensionality, instructional methods, response patterns.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

A computerized diagnostic adaptive test for a series of pre-algebra signed-number lesson (which are also on the computer system) was programmed along with a computer-managed routing system by which each examinee was sent to the instructional unit corresponding to the level of skill at which she/he stopped in the initial test. Upon completion of the course a computerized conventional posttest was given to the examinees. The posttest scores were far from being unidimensional, while

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

the pretest and posttest data obtained from a previous study, in which the pretest was a computerized conventional test and students were forced to go through all instructional units regardless of their achievement in the pretest, indicated a strong tendency to be unidimensional. The response patterns of the posttest in the present study showed a high error rate for the skills prior to stopping levels for a subgroup of examinees.

A cluster analysis was performed on the response patterns and four different groups were found. A discriminant analysis indicated significant differences among the four groups in response patterns of the skills in signed number operations. After interviewing the teachers and several children, we came to the conclusion that it was the difference between prior and current instructional methods that confused students and caused a mess in the posttest data. In other words, there was a proactive inhibition effect.

The scoring procedure of the adaptive testing did not consider individual differences in information-processing skills which were affected by the instructional method used in previous teaching. Thus, the students who were taught to perform the beginning part of a set of hierarchically ordered skills by instructional method A would very likely get confused in a lesson in which a different instructional method B was adopted. Consequently, quite a few number of peculiar response patterns were seen in the performance on the posttest. This fact cautions us that one should be careful not to rely solely on test results determined by performance scores on a diagnostic pretest when a computer-managed instructional system is to route each examinee to their level of instruction. It was suggested that we must somehow unravel what information-processing strategy was used and consider this knowledge simultaneously.

THE DANGER OF RELYING SOLELY ON DIAGNOSTIC ADAPTIVE TESTING WHEN PRIOR AND SUBSEQUENT INSTRUCTIONAL METHODS ARE DIFFERENT

by

Kikumi Tatsuoka
and
Menucha Birenbaum

University of Illinois at Urbana-Champaign

Approved for public release; distribution unlimited.
Reproduction in whole or in part permitted for any
purpose of the United States Government.

This research was sponsored by the Personnel and Training Research Programs, Psychological Science Division, Office of Naval Research, under Contract No. N00014-78-c-0159. Contract Authority Identification Number, NR 150-415.

Research Report NO. 2

March 1979

ACKNOWLEDGEMENT

We wish to acknowledge the kind cooperation extended us by the mathematics teachers at Urbana Junior High School, particularly Ms. Mary Klein.

Julie Garrard gave us helpful editorial assistance, and Wayne Wilson provided careful artwork.

THE DANGER OF RELYING SOLELY ON DIAGNOSTIC ADAPTIVE TESTING WHEN PRIOR AND SUBSEQUENT INSTRUCTIONAL METHODS ARE DIFFERENT

Kikumi Tatsuoka

and

Menucha Birenbaum

ABSTRACT

A computerized diagnostic adaptive test for a series of pre-algebra signed-number lessons (which are also on the computer system) was programmed along with a computer-managed routing system by which each examinee was sent to the instructional unit corresponding to the level of skill at which she/he stopped in the initial test. Upon completion of the course a computerized conventional posttest was given to the examinees. The post-test scores were far from being unidimensional, while the pretest and post-test data obtained from a previous study, in which the pretest was a computerized conventional test and students were forced to go through all instructional units regardless of their achievement in the pretest, indicated a strong tendency to be unidimensional. The response patterns of the post-test in the present study showed a high error rate for the skills prior to stopping levels for one subgroup of examinees.

A cluster analysis was performed on the response patterns of the skills and four different groups were found. A discriminant analysis indicated significant differences among the four groups in

response patterns of the skills in signed number operations. After interviewing the teachers and several children, we came to the conclusion that it was the difference between prior and current instructional methods that confused students and caused a mess in the post-test data. In other words, there was a proactive inhibition effect.

The scoring procedure of the adaptive testing did not consider individual differences in information processing skills which were affected by the instructional method used in previous teaching. Thus, the students who were taught to perform the beginning part of a set of hierarchically ordered skills by instructional method A would very likely get confused in a lesson in which a different instructional method B was adopted. Consequently, quite a few peculiar response patterns were seen in the performance on the post-test. This fact cautions us that one should be careful not to rely solely on test results determined by performance scores on a diagnostic pretest when a computer-managed instructional system is to route each examinee to their initial level of instruction. It was suggested that we must somehow unravel what information processing strategy was used and consider this knowledge simultaneously.

INTRODUCTION

The computer-based education system (PLATO) at the University of Illinos has been widely utilized in teaching many different subject areas. The mathematics program at Urbana Junior High School (UJHS) is one of many that are currently on the PLATO system. Four terminals have been installed in the Mathematics Laboratory at UJHS so that they would be used by students from different classes as a part of their mathematics curriculum, and they had about an 80% rate of utilization during the time in operation. An increasing number of teachers have shown their interest in being involved with the PLATO mathematics program each semester. A great majority of the students, ranging from the best to the worst seemed to enjoy working with the PLATO lessons, especially with the game-lessons (Weaver, 1978).

About 70 lessons that teach a wide variety of subject areas from fundamental arithmetic such as decimal numbers to algebra and geometry have been presented by the system router, which allows a student or a teacher to choose a lesson from the index of available materials. This freedom of choice could be a troublesome task for a teacher because she has to determine which lesson would be the most appropriate instructional material for students who need remedial study of some topics. Moreover, without a larger number of terminals available no greater amount of time could be available for a student. Thus, adaptive diagnostic testing and computerized routing systems based on the results of the initial test become essential.

An adaptive test of signed numbers consisting of 12 groups of

items which represent 12 different skills was implemented along with a computer managed routing system. About 120 students took the initial test of adaptive testing although only 92 students completed the computerized conventional post-test given at the end of the instruction. It seemed that the children liked this "strange" format of testing. Some of them even volunteered to try the test for fun. However, the response patterns of the post-test revealed that the error-rate of the skills prior to the examinee's stopping level at the initial test was disturbingly high for some students. This fact contradicted our expectation that the scores on the post-test would satisfy a sufficient condition of the assumption of local independence--i.e., unidimensionality.

A close investigation of the behavior of the response patterns led us to consider a new aspect of the scoring procedure in adaptive testing which has been traditionally neglected.

A cluster analysis was performed on the 92 examinees' response patterns on the basis of Euclidian distances between pairs of response vectors, and four different groups were found. A discriminant analysis indicated significant differences among the four groups in terms of total scores on the 12 skills. After interviewing the teachers and several children, we came to the conclusion that it was the difference between prior and current instructional methods that confused the students and caused a mess in the post-test data. The two conflicting instructional methods created difficulty in following the instructions in the PLATO lesson for those students who operated addition of signed numbers by the method taught by one of the teachers. As will be

discussed later, the procedures of information processing associated with these two instructional methods of performing arithmetic upon signed numbers are greatly different. The traditional scoring procedure of the latent trait theory would not be capable of detecting these discrepancies based on the different information processes for arriving at the answers to a given item.

METHOD AND PROCEDURES

Pretest: A computerized conventional pre-test consisting of 64 items among which 4 or 6 items represented each of 14 different skills of integer (or signed number) operations was given to the pre-algebra classes at UJHS during the Spring semester of 1978. These items were displayed on the PLATO screen one at a time and were kept there until the student typed in his/her answer. No feedback, including a simple judging of either OK or NO to the answer, was given during the testing. After the pretest was taken, the classes began studying signed number operations, and at the same time the PLATO lessons started. The student eventually completed all instructional units in the lessons in which 14 skills were taught. Since the contents of these lessons adopted different teaching methods which became a crucial theme of this study, a brief description of the methods will be given below.

The Number Line Method: In the lesson "signum" written by Tamar Weaver, the concept of negative integers was taught by means of moving a pointer to the left on the number line starting from the origin zero. Addition

of numbers was associated with moving a pointer to the right by the number of units equivalent to the addend, while subtraction was taught by moving the pointer to the left by the number of units corresponding to the subtrahend. This geometric method did not seem to be successful in teaching problems involving double signs. (Problems such as: (-1)-(-7) as Weaver [1978] pointed out in his paper.) Students seemed to have trouble in understanding how double signs work in a geometric way, with a negative sign in front of a negative number causing a pointer to be reflected through the origin on the number line. Students who were successful in problems with double signs showed a different way of approaching the problem.

Madison Mathematics Project (Davis, 1964): A new approach was presented by this project. Positive and negative integers were associated with checks and bills, respectively. Addition was represented by a mailman's bringing something (a check or a bill), while subtraction corresponded to the mailman's taking something from the house. Also this method didn't use parentheses such as 1+(-3) which appears in the Number Line method. Instead, signs were written at the upper left of a number, like $1^{+}\ {}^{-}3$ and it was clearly distinguished from an operational sign of addition + . With this approach students were very successful at working problems involving double signs but many failed to see the problem as a signed number subtraction problem in general. Weaver stated that when the students using this method were asked to work the problem as a subtraction problem directly, they failed.

The teachers' method: A teacher who is the head of the mathematics department at UJHS followed the Madison project method but he has changed the presentation of problems substantially. He has used a sequential method: that is, starting with a problem that children already know, say 5 - 3 = 2 he subsequently presented 5 - 4 = 1, 5 - 5 = 0. Next, he asked what 5 - 6 will be, assuming children know that -1 lies immediately to the left of zero on the number line.

In contrast to the above, another teacher taught signed number operations mainly by showing examples of various skills and emphasized memorizing the rules of operations.

Treatment and Post-test: The class of Spring 1978 studied two types of lessons: one in which one method was used to teach signed-number operations, and the other, the other method. A computerized conventional 64-item test was administered upon completion of the PLATO course. This post-test will be referred to as post-test1 hereafter in the paper.

Diagnostic Adaptive Testing: An unconventional test was constructed on the bases of item characteristic information obtained from the pretest item scores.

First, the computer program that estimates item discriminating powers (a's) and item difficulties (b's) of the two parameter logistic model, by the maximum likelihood method, was written on the PLATO system by Robert Baillie. The iterations in the program (called "getab") start

with a special set of initial parameter values (given by Lord & Novick, 1968, Chapter 16) and continues until the estimated parameter values converge to the constant. This program successfully provided the convergent estimated values of a's and b's for most of the 64 items in the pre-test and θ values for 83 students. Since the pretest was a computerized free-response test, it was considered that guessing would be a negligible factor. Moreover, the coefficient α of the pretest was as high as .974.

The adaptive test of signed numbers consisted of 12 groups of items representing 12 different skills. The pretest contained 14 different skills but two skills were dropped in the diagnostic adaptive test, due to the shortage of PLATO terminals available at UJHS. Besides that, the lessons that dealt with those skills, multiplication and division of signed numbers, would have added another 50-60 minutes to the program.

In the pretest, one item from each skill of the total of 14 skills was given first, then the second item from all 14 skills was given in the same order of skills as in the first 14 items. Thus, each skill was examined by either four or six parallel items in the test. A close examination of the items testing the same skill revealed that the item parameters of these parallel items did not show much noticeable difference. Therefore, the averages of a's and b's in a skill were taken to designate the characteristics of each of the 14 skills. Table 1 presents examples of items from each of the 12 skills that were used in the adaptive test and their average indices of item difficulties, means and standard deviations.

Table 1

Means and Item Difficulties of the Twelve Skills in the Pretest of Signed Numbers in Spring 1978

	Skill Type	Mean	SD	Difficulty
1	2>-5	2.66	1.69	-1.67
2	6 right of -2	1.87	1.76	-1.51
3	(-2)+(-5)	3.15	1.28	-1.37
4	-8+7	3.00	1.51	-1.06
5	2+(-6)	3.21	1.19	-.86
6	(-3)-(-2)	1.51	1.74	-.76
7	5-6	1.71	1.84	-.49
8	-(-7)	1.79	1.73	-.01
9	(-4)-(-6)	2.61	1.55	.10
10	-1-5	3.65	2.32	.15
11	2-(-7)	3.50	2.25	.22
12	(-6)-(+5)	5.22	1.31	.36

Procedure of Adaptive Testing: The newly developed adaptive test and routing system were tried during the Fall semester of 1979. Administering the test to the classes of 8th graders began a week after the regular classroom instruction started teaching the signed-number operations. All students were expected to know the number line and what negative integers are. Moreover, most students had learned more or less how to add any two integers. Thus, the starting item of the diagnostic adaptive test was selected from Skill No. 6, (-3)-(-2) type. According to the result of his/her answer, a skill either one step harder or one step easier was next tested. This procedure was repeated until the "stopping criterion" was satisfied. Each examinee was routed to the instructional unit corresponding to the level of skill at which he/she stopped in the initial test. The instructional units of the PLATO lesson that teach the same 12 skills by the Number Line method also were rearranged into the same order as the skills in the adaptive test, so that if an examinee stopped at the 7th skill level he was sent to the

7th instructional unit. After he went through the 7th to 12th instructional units, the student completed the lesson and a 52-item conventional computerized post-test (post-test2) was administered to him. Performance score and response latency of each item as well as the performance records and mastery time of each instructional unit were collected for all students.

Stopping Criterion of Testing: The θ values estimated by the maximum likelihood method were not used in routing students into the lesson because this estimation method did not converge for all response patterns. Nonconverging cases will halt the routing system thus forcing us to forgo using the θ values to decide on the stopping levels of the skills. If the estimated θ values are always obtainable, then it is commendable to choose a subsequent item from the remaining items so as to maximize the amount of information at a subject's true ability level θ (Samejima, 1978). Tatsuoka (1979) derived the least-squares estimation method of the θ values by a Hilbert space approach. The beta weights for earlier terms in the multiple regression equation remain unaltered when subsequent terms are added in a stepwise manner. With this method, the θ values are always obtainable even for unusual response patterns or extreme values of θs as well. This method will be applied to the future use of adaptive testing on the PLATO system. Our current stopping criterion is similar to the one of stradaptive testing which was discussed by Weiss (1973), Dewitt and Weiss (1974), and Waters (1978).

ANALYSES AND CLASSIFICATION OF RESPONSE PATTERNS

Dimensionality of Post-test2: We assumed first that the post-test2 data would not have obviously more than two dimensions, inferring from the fact that both the pretest and post-test1 data had a strong tendency toward unidimensionality. Therefore, we arranged the skills into a linearly related hierarchical structure and applied the latent trait model to determine item characteristic parameters. Figure 1 presents a scree-test of the eigenvalues obtained by a principal component analysis for the pretest, post-test1 and post-test2 data. As can be seen in the figure, the pre and post-test1 data share an almost identical pattern, which consists of one substantial eigenvalue that accounts for about half of the variance (Table 2 presents the amount of variance accounted for by each eigenvalue). The second eigenvalue, the magnitude of which exceeds a unity, accounts for only 13% of the variance. The pattern of the post-test2 data is different. Four eigenvalues in this case exceed

Table 2

The Percent of Variance Accounted for by the Corresponding Eigenvalues in a Principal Component Analysis of Pre and Post Data of the Classes of Spring 1978 and the Posttest of Adaptive Test Study in Fall of 1978

	Spring-1978 Percent of Variances		Fall-1978 Percent of Variances
	Pretest	Posttest	Posttest
1	53.1	49.8	27.6
2	13.6	13.0	14.9
3	6.3	7.7	12.9
4	5.7	6.1	9.2
5	4.2	5.2	7.0
6	3.9	3.9	6.2
7	3.0	3.3	5.7
8	2.8	2.5	5.3
9	2.3	2.3	3.8
10	1.3	2.0	3.3
11	1.3	1.5	2.6
12	1.1	1.4	1.6
13	.7	.7	
14	.6	.6	

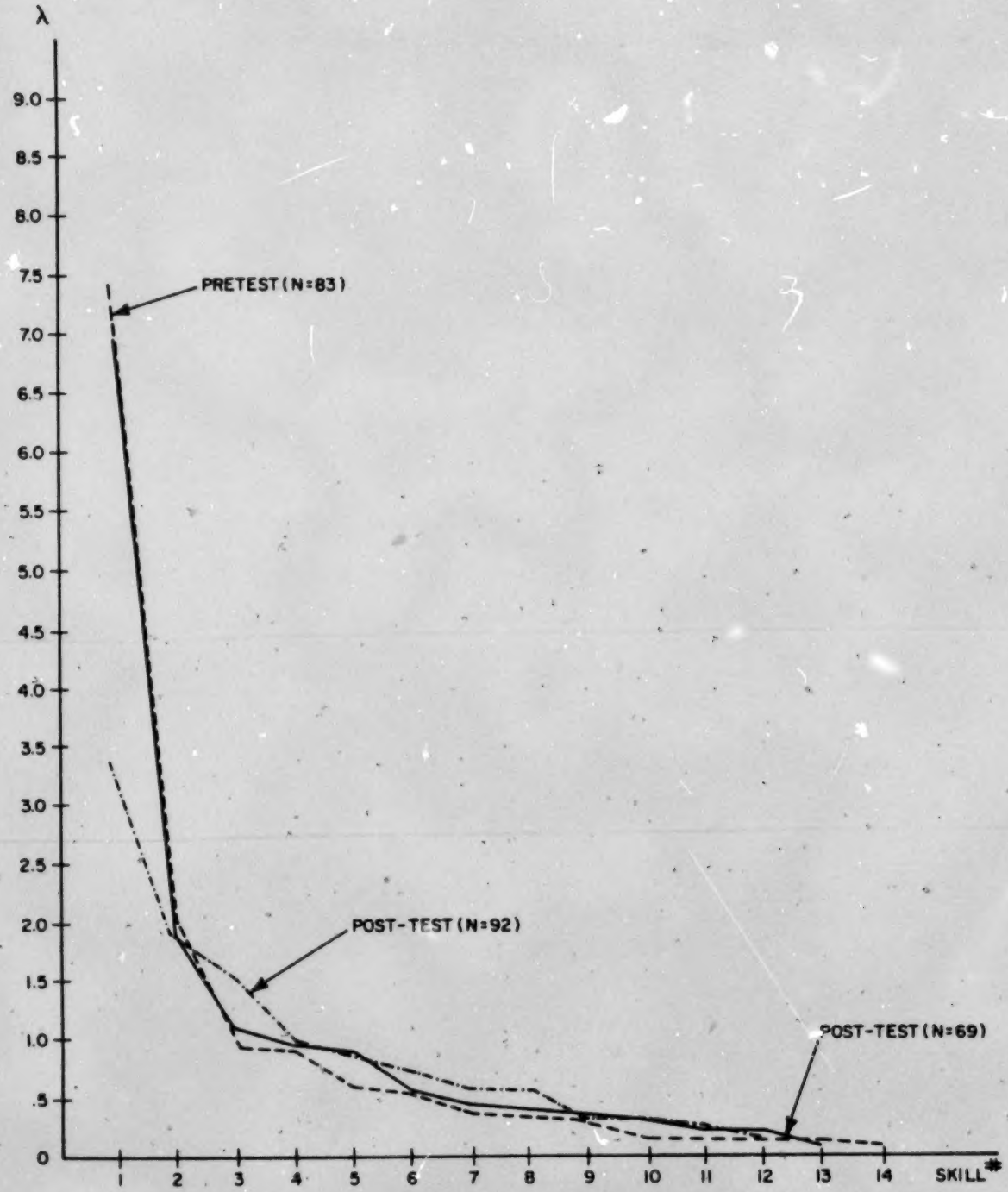


Figure 1 Screetest: Eigenvalues extracted in a principal component analysis.

unity but the differences in magnitude among these four eigenvalues are relatively smaller. The amount of variance accounted for by all four eigenvalues is 65%. We can therefore conclude that while the post-test1 data shows a certain tendency toward unidimensionality, the structure of the post-test2 data departs from unidimensionality to a much greater extent.

The increasing dimensionality of the post-test data as compared with those of the previous group may indicate that other factors were interacting with the basic ability of manipulation with signed numbers. As was mentioned before, in the adaptive testing program students were taught the basic levels of signed numbers by means of the ordinary classroom teaching methods which in many cases happened to differ from the method presented by the computerized instruction. These differences referred not only to the medium of presentation, the style and the notation used, but more importantly they differed also with respect to the conceptualization of the material. We can therefore assume that for some students in this group, the adaptive instruction--to which they were routed according to their performance in the pretest--was a different experience from what they were previously taught. This situation in which the teaching method wasn't consistent with their former background not only caused their lack of understanding for the new material, but may also have confused them as to materials they had previously mastered. This, of course is one possible explanation.

An alternative explanation may question the reliability of the routing procedure, claiming an invalid hierarchy or violation of the

local independence assumption underlied the logistic model, upon which the routing process was based. Although the pretest data showed a tendency toward unidimensionality which is a prerequisite for hierarchical structure as well as for the latent trait model, there were still other sources of systematic variation in the data. This may have caused less reliable estimates regarding the starting point in the instructional unit.

Although the first explanation seems more plausible, it is impossible at this stage to exclude the second one entirely. In order to do so, an experimental design should have been carried out including a third group which should have taken the adaptive test and the entire instructional unit, and a fourth group which should have been taught the previous material in a method similar to the one offered by the computerized instructional unit. A comparison of the groups' results in the post-test may have lent support to one of the above mentioned tentative explanations.

Identifying a Typology of Response Patterns on Post-test2: The multidimensionality that emerged in the post-test2 data of the adaptive testing group indicated the existence of different patterns of responses to the 12 skills measured by that test. In order to classify the different patterns into a more meaningful typology, a cluster analysis was applied (i.e., students were clustered according to the similarity of their responses on the post-test items). The method of clustering the cases was based on the hierarchical model (Hubert & Baker, 1976). The computer program used was the BMDP2M (1977). This

procedure performs an hierarchical cluster analysis based on the average linkage algorithm. Initially the program considers each object to be in a cluster of its own. At each step the two clusters with the shortest distance (which is defined by Euclidian distance between two response vectors) between them are combined and treated as one cluster. This process of combining clusters continues until all the objects are combined into one cluster. The final result obtained on letting the computer program run its full course is obviously a trivial one, for it constitutes no partitioning of the original total group. A partitioning at some intermediate stage must be chosen on the basis of some criterion involving "a trade-off between the loss of information as the partition level increases (i.e., as larger groups are formed) and the greater ease with which substantive interpretations made by the research when the number of groups in the partition is small" (Hubert & Baker, 1976). In the present case, a careful examination of the tree diagram printed by the computer program led to a partition with four subgroups of students, defined by their response pattern to the 52 post-test items. In order to validate this classification and to identify the differences among the four response-pattern types in terms of the 12 skills, a discriminant analysis was carried out. Tables 3 and 4 present the results of this analysis. As can be seen in the tables, two functions yield highly significant discrimination among the groups. The discriminant analysis enables us to identify the nature of these response types.

Table 3

Centroid & Coefficient of Determination for the 12 Skills on the 3 Discriminant Functions (N=91)

Group	N	I	II	III
1	34	-.571	.280	-.280
2	27	.951	-.132	-.160
3	20	-.005	.427	.590
4	10	-.614	-1.448	.203
λ		.775	.446	.135
R_c		.661	.555	.345
$(1-\Lambda)100$		66%	39%	12%
p<		.0001	.001	.230

Table 4

Standardized Coefficient for the 12 Skills on 3 Discrininant Functions

	Discriminant Functions		
Skill	I	II	III
1	.397	.186	-.627
2	.479	-.038	.601
3	.175	.534	-.196
5	.185	-.209	.350
6	-.109	.773	-.055
7	.289	.180	-.243
8	-.244	.313	-.260
9	-.214	-.468	.670
10	-.150	.401	.725
11	-.517	.082	-.622

One dimension along which the greatest differences occur involves skills 1,2,7 vs. 8,9,11. This dimension best discriminates Group 2 from the rest. As can be seen in Figure 2, the profile of this group when considered for each skill separately reflects an extreme profile. While this group compared with the other three showed the best performance on the lower level skills, its performance on the higher

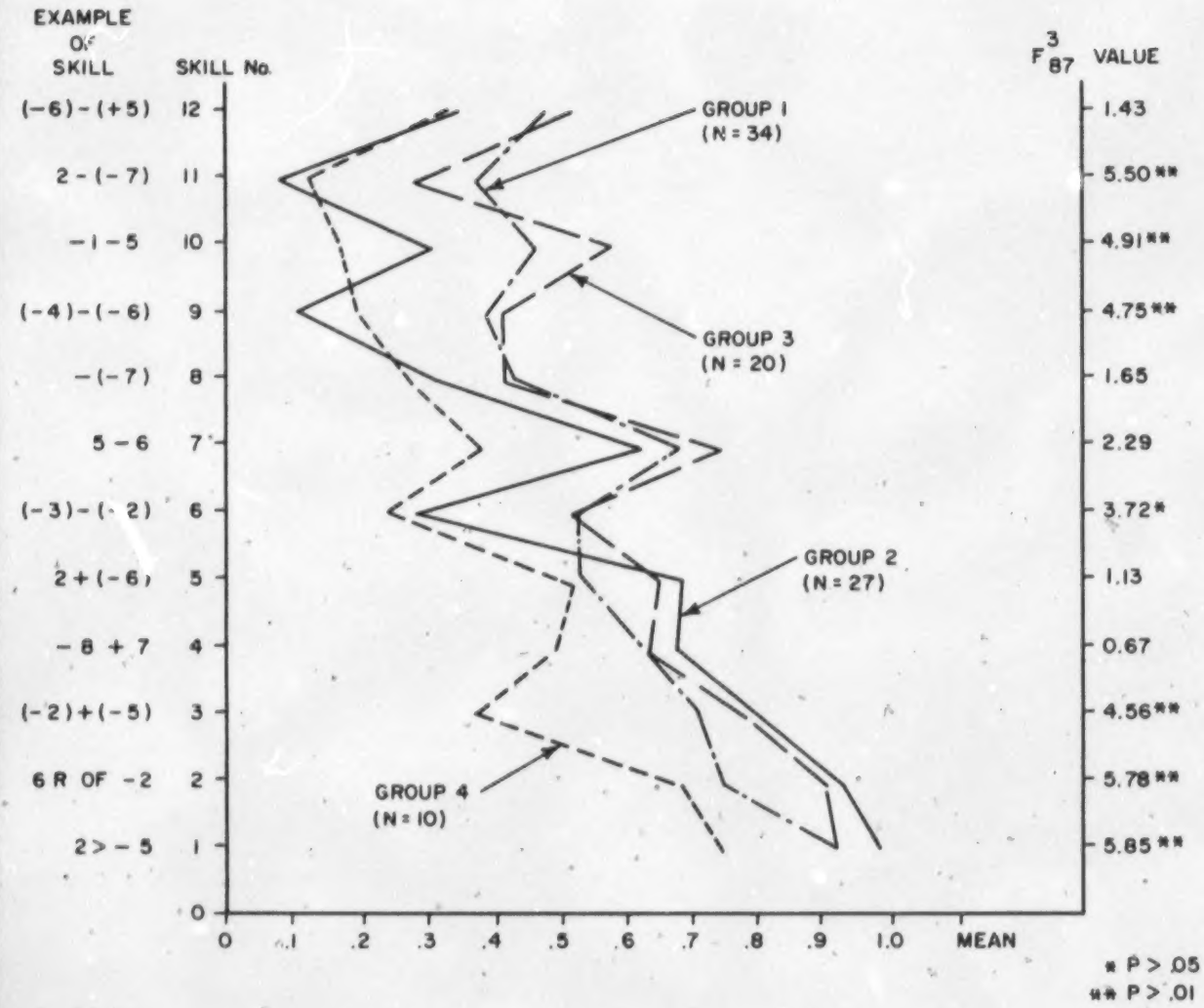


Figure 2 Profiles of the mean performance on each skill in the post-test for four groups formed via cluster analysis.

level skills is almost the poorest. According to the information supplied by the classroom teachers regarding the students' backgrounds, most of the students in this group were taught previously by the sequential method closely associated with the Madison Program. As was mentioned before, this instructional method differs from that of the number line with respect to the conceptualization of the integer operations.

The second function discriminates Group 4 from the rest of the groups. As can be seen in Figure 2 this group shows a poor performance along the entire test. The information supplied by the classroom teachers confirmed that most of the students that have been clustered to this group are of low math ability as can be judged from their math grades during the previous year.

Reliability of the Classifications: Table 5 presents the actual and predicted group affiliations. As can be seen in the table, the overall rate of correct classification is 65%. For Group 2 the prediction was most accurate, resulting in the highest rate of 89% of the cases in this group being correctly classified on the basis of the discriminant function scores.

Based on this experience it seems that a cluster analysis of the response patterns followed by a discriminant analysis has the potential of providing valuable information that may help to identify problems in the teaching-learning process.

Table 5

Actual and Predicted Group Afiliation of
the Four Response Pattern Types (in percent)

		Predicted Groups				
Actual Groups	N	1	2	3	4	%total
1	34	55.9	0.0	26.5	17.6	37.4
2	27	3.7	88.9	7.4	0.0	29.6
3	20	20.0	30.0	45.0	5.0	22.0
4	10	0.0	30.0	0.0	70.0	11.0
Predicted %		26.4	36.3	22.0	15.3	100.0

64.8% of the cases were correctly classified.

DISCUSSION AND CONCLUSIONS

The results of this study raised two important, albeit closely related, issues concerning adaptive testing and computer managed routing by which each examinee was sent to his/her most appropriate instructional level, i.e., his/her adaptive instructional unit, diagnosed by the initial adaptive test.

The first issue is how one could improve the scoring procedure of adaptive testing by taking into account individual differences in information processing skills that were usually affected by instructional method used in previous teaching. Many psychological studies pertinent to an information processing view of mental abilities have been done recently by cognitive psychologists. (See for example: Anderson et al., 1978; Carroll, 1978; Frederiksen, C., 1969; Frederiksen, J., 1978; Groen & Perkum, 1972; Heller & Greeno, 1978; Hunt et al., 1973; Rose, 1977; Sternberg, 1978a, 1978b; Sternberg & Rifkin, 1978.) The results of these

studies have indicated the exsistence of a series of cognitive processes which differed among individuals. However, the stability and the generality of these traits have not been definitely confirmed yet.

A cluster analysis performed on the similarity of response patterns on the test separated a group whose members seemed to use alternative processes in doing some test items. They did very well on the items of the skills prior to the stopping level on the diagnostic adaptive test, but did not learn much in their adaptive instructional units. When the list of names in this group was presented to the teachers, it caused surprise because the members of this group were considered fairly good students and the teachers expected them to be able to perform much better on the test. The two conflicting instructional methods, a prior instructional method taught by a teacher and a subsequent one presented by the adaptive instructional unit caused confusion in learning non-mastered skills for those students. It therefore seems that in order to improve the adaptive procedure, the students' strategy of information processing, due to their previous learning experience should be taken into consideration as well.

The second issue was the dimensionality of the performance scores on the test administered at the end of adaptive instructions. Since the items in post-test1 and post-test2 are identical, we expected that the dimensionality of post-test2 would be almost the same as that of post-test1. Post-test1 data obtained from a previous study showed a strong tendency toward unidimensionality, but post-test2 data did not show it. In order to speculate about the reason why the dimensionalities of the two post-tests are different, we must examine

the differences between the amounts of treatment given in the previous study (Spring of 1978), and the current study (Fall of 1978). In the previous study, the students studied both PLATO lessons: One in which the Madison Project approach was used, and the other, the Number Line approach to teach signed-number operations. Moreover, the students represented in post-test1 data studied a whole segment of the lesson by the Number Line Method, while those in post-test2 data studied only a part of this lesson--starting from the unit to which they were routed. Naturally, the means of all skills in post-test1 were higher than those in post-test2, and 7 out of 12 skills were significantly high. This discrepancy may imply that if a given topic is considerably well mastered by a majority of students, the post-test will show a strong tendency toward unidimensionality no matter what kinds of information-processing strategies were used by individuals. As was seen in this study, on the other hand, when learning is still far from the stage of mastery, and different instructional methods create confusion among students, the dimensions of a test given at this point will be chaotic.

Solutions to the first issue have not been fully explored. However, it seems that multivariate assessment of performance scores and response latency, or more precisely, the notion of conditional response rate (Tatsuoka & Tatsuoka, 1978), might be an appropriate solution. Exploration of time data indicated that the curves of conditional response rate function (or hazard rate function [Mann et al., 1974]) of a given item in Group 1 and 2 found in the cluster analysis results were obviously different from one another. That is, the curve obtained from Group 2 strongly suggested a Poisson process for the problems of

additions, while that from Group 1 had a monotonically increasing conditional response rate function for the same items. This matter will be further discussed in Technical Report No. 3.

A second possible solution to the first issue is to investigate closely the behavior of response patterns on the test. As mentioned earlier, each instructional method had a unique strength and weakness in teaching a given skill. It was easier to teach problems of double signed numbers with the Madison Project method than with the Number Line method. Application of S-P curves (Sato, 1977; K. Tatsuoka, 1978; M. Tatsuoka, 1978) or Cliff's consistency index (1977) seem to have the potential of providing a solution to the above mentioned issue of improving the scoring procedure for adaptive testing.

REFERENCES

Anderson, J.R., Kline, P.J., Beasley, C.M., Jr. A theory of the acquisition of cognitive skills. New Haven: Yale University, Department of Psychology, 1978.

Carroll, J.B. How shall we study individual differences in cognitive ability? -- Methodological and theoretical perspectives (Technical Report 1). Chapel Hill: University of North Carolina, Psychometric Laboratory, 1978.

Cliff, N.A. A theory of consistency of ordering generalizable to tailored testing. Psychometrika, 1977, 375-399.

Davis, R.B. The Madison Project, discovery in mathematics; a text for teachers. Reading: Addison-Wesley, 1964.

DeWitt, L.J., & Weiss, D.J. A computer software system for adaptive ability measurement (Research Report 74-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Laboratory, 1974. (AD 773961)

Frederiksen, C.H. Abilities, transfer, and information retrieval in verbal learning. Multivariate Behavioral Research Monographs, 1969, No. 69-2.

Frederiksen, J.R. Assessment of perceptual decoding and lexical skills and their relation to reading proficiancy. In A.M. Lesgold, J.W. Pellegrino, S. Fokkema, and R. Glaser (Eds.), Cognitive psychology and instruction. New York: Plenum Press, 1978.

Groen, G.J., & Perkum, J.M. A chronometric analysis of simple addition. Psychological Review, 1972, 79, 329-343.

Heller, J.I., & Greeno, J.G. Information processing analyses of mathematical problem solving. Unpublished manuscript, University of Pittsburgh, 1978.

Hubert, L., & Baker, F.B. Data analysis by single-link and complete-link hierechical clustering. Journal of Educational Statistics, 1976, 1(2), 87-111.

Hunt, E., Frost, N., & Lunneborg, C. Individual differences in cognition. In G. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol.7). New York: Academic Press, 1973.

Lord, F.M., & Novick, M.R. Statistical theories of mental test scores. Reading: Adison-Wesley, 1968.

Mann, N.R., Schafer, R.E., & Singpurwalla, N.D. Methods for Statistical analysis of reliability and life data. New York: John Wiley & Sons, 1974.

Rose, A.M. An information processing approach to performance assessment (Final Report). Washington, D.C.: American Institutes for Research, November 1978.

Samejima, F. Effects of individual optimization in setting the boundaries of dichotomous items on accuracy of estimation. Applied Psychological Measurement, 1978, 1, 77-94.

Sato, T. Diagnostic and formative evaluation of data processing system. In T. Sato (Ed.), Computer managed instruction system. Tokyo: Electronics Communication Inc., 1977.

Sternberg, R. The nature of mental abilities (Technical Report No. 9). New Haven: Yale University, Department of Psychology, 1978a.

Sternberg, R. Contrasting conceptions of intelligence and their educational implications (Technical Report No. 14). New Haven: Yale University, Department of Psychology, 1978b.

Sternberg, R., & Rifkin, B. The development of analogical reasoning processes. Journal of Experimental Child Psychology, in press, 1978.

Tatsuoka, K.K. Psychometrical and statistical properties of response time. Paper presented at the joint meeting of the Psychometric Society and the Society for Mathematical Psychology, Hamilton, Ontario, June 1978.

Tatsuoka, K.K. The least-squares estimation of latent variables by a Hilbert space approach (Research Report E-3). Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1979.

Tatsuoka, K.K., & Tatsuoka, M.M. Time-score analysis in criterion-referenced tests (CERL Report E-1). Final Report for NIE project No. NIE-G-76-0087. Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, February 1978.

Tatsuoka, M.M. Recent psychometric developments in Japan: Engineers grapple with educational measurement problems. Paper presented at ONR Contractor's meeting, Columbia, Missouri, September 1978.

Waters, B.K. An empirical investigation of the stratified adaptive computerized testing model. Applied Psychological Measurement, 1978, 1, 141-154.

Weaver, C. Classroom observation report: Urbana Junior High School PLATO project. Unpublished manuscript, University of Illinois, Computer-based Education Research Laboratory, 1978.

Weiss, D.J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Laboratory, 1973. (AD 768376)